DESIGNING AN ALGORITHM TO PREDICT
THE INTENSITY OF THE SEVERE
WEATHER SEASON

THESIS

Hugh J. Freestrom, First Lieutenant, USAF

AFIT/GM/ENP/02M-04

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
14 Jan 02	Final	Jun 01 - Mar 02

Title and Subtitle Designing an Algorithm to Predict the Intensity of the Severe Weather Season	**Contract Number** **Grant Number** **Program Element Number**
Author(s) 1stLt Hugh J. Freestrom, USAF	**Project Number** **Task Number** **Work Unit Number**
Performing Organization Name(s) and Address(es) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Bldg 640 WPAFB OH 45433-7765	**Performing Organization Report Number** AFIT/GM/ENP/02M-04
Sponsoring/Monitoring Agency Name(s) and Address(es) AFCCC/DOC5 Maj Joseph King 151 Patton Ave., Rm 120 Asheville, NC 28801-5002	**Sponsor/Monitor's Acronym(s)** **Sponsor/Monitor's Report Number(s)**

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
The original document contains color images.

Abstract
Examination of atmospheric and oceanic circulations may explain interannual climate variability in the Northern Hemisphere on a seasonal scale. It is crucial to develop more accurate seasonal climate forecasts using both global circulation and sea surface temperature (SST) indices to aid in long-range weather forecasts. These global circulation and SST indices are becoming increasingly available to worldwide users and using them for seasonal prediction has spread not only to scientists, but also to brokerage firms, utilities, and the Department of Defense (DoD). DoD is extremely interested in long-range seasonal forecasts of severe weather for asset protection, mission planning, and worldwide operations. The goal of this research was to create a predictive algorithm for locations in the southeastern and south-central portion of the United States in support of the Air Force Combat Climatology Center (AFCCC) to use in predicting the intensity of the spring and summer severe weather seasons.

Subject Terms Atmospheric circulations, Long-range forecast, Multiple linear regression, Classification trees, Sea-surface temperatures, Algorithms, Severe weather	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 70	

AFIT/GM/ENP/02M-04

DESIGNING AN ALGORITHM TO PREDICT THE INTENSITY OF THE SEVERE WEATHER SEASON

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Meteorology

Hugh J. Freestrom, B.S.

First Lieutenant, USAF

March 2002

AFIT/GM/ENP/02M-04

DESIGNING AN ALGORITHM TO PREDICT THE INTENSITY

OF THE SEVERE WEATHER SEASON

Hugh J. Freestrom, B.S.
First Lieutenant, USAF

Approved:

Ronald P. Lowther (Chairperson)	Date
Gary R. Huffines (Member)	Date
Edward D. White (Member)	Date

Acknowledgements

I would like to thank many people who made this thesis work possible. First, I would like to thank my thesis advisor, Lt Col Ronald Lowther, for his superfluous technical assistance and also his mentorship during this process. Second, I would like to thank all other members in my committee, Maj Gary Huffines and Maj Edward White, for their expertise they provided in their fields of study. Additionally, I would like to thank my colleagues that all contributed scientific advise into this thesis. Finally, I would like to thank my friends and family who were patient and understanding during the time spent on this work.

Hugh J. Freestrom

Table of Contents

List of Figures

Figure

List of Tables

Abstract

Examination of atmospheric and oceanic circulations may explain interannual climate variability in the Northern Hemisphere on a seasonal scale. It is crucial to develop more accurate seasonal climate forecasts using both global circulation and sea surface temperature (SST) indices to aid in long-range weather forecasts. These global circulation and SST indices are becoming increasingly available to worldwide users and using them for seasonal prediction has spread not only to scientists, but also to brokerage firms, utilities, and the Department of Defense (DoD). DoD is extremely interested in long-range seasonal forecasts of severe weather for asset protection, mission planning, and worldwide operations. The goal of this research was to create a predictive algorithm for locations in the southeastern and south-central portion of the United States in support of the Air Force Combat Climatology Center (AFCCC) to use in predicting the intensity of the spring and summer severe weather seasons.

The most significant predictor of the intensity of the severe weather season in the southeast and south-central regions of the U.S. was identified as the proximity of the indices to the respective region. Beginning with multiple linear regression, this study found there were relationships between several severe weather parameters, such as thunderstorm and heavy precipitation events, and these known global circulation and SST indices. However, R^2 values showed that SST indices had more significance with severe weather since they appeared more often in the multiple linear regression models. In addition, analysis of variance provided valuable incite into the development of

classification and regression tree (CART) analysis. After little predictive value was found using traditional statistics, CART analyses were developed to create an algorithm for DoD forecasters to use for seasonal severe weather prediction. Results confirmed that algorithms with reasonable predictability can be produced for forecasting the intensity of the severe weather season.

DESIGNING AN ALGORITHM TO PREDICT THE INTENSITY

OF THE SEVERE WEATHER SEASON

I. Introduction

One of the greatest challenges in meteorology today is long-range forecasting. Weather-sensitive industries such as agriculture and energy use long-range climate forecasts to project future crop yields and the amount of natural gas or electricity required for a season. The Department of Defense (DoD) is also extremely in need of these forecasts. DoD is responsible for examining the influences of long-term weather phenomena on its operations by using future seasonal outlooks, especially for severe weather phenomena.

Operational commanders routinely task the Air Force Combat Climatology Center (AFCCC) to produce outlooks for the upcoming severe weather season so they can tailor their operations to meet any threat. One possible use of such forecasts in the United States is the realignment of aircraft to optimize their training and operational effectiveness. However, at the present time, AFCCC does not have the capability to produce such outlooks. The goal of this research therefore, is to develop a predictive algorithm for the southeastern and south-central portion of the United States in support of AFCCC to use in forecasting the intensity of the spring and summer severe weather seasons.

1.1 Statement of the Problem

Sea-surface temperatures (SSTs) are superb indicators that climatologists and weather sensitive groups use for long-range forecasts since they are known to control some of the interannual climate variability in all regions of the globe. Since the oceans cover nearly 70 percent of the Earth's surface, absorbing and reradiating enormous amounts of solar radiation, SST patterns driven by ocean currents greatly affect the character of weather patterns downstream, particularly across North America (Sanders, 1985). Interest in SSTs, such as in the Pacific Basin, the North Pacific, and the North Atlantic, has spread not only to scientists, but also to primary agricultural producers, brokerage firms, and the military. Although it is difficult to explain every aspect of SSTs and their influences globally, relationships exist between them and with temperature, precipitation, and severe weather anomalies throughout the United States.

Another indictor scientists use are the global atmospheric circulation patterns. For example, one of the most influential known global atmospheric circulations is associated with the Pacific Basin and its associated El Nino/Southern Oscillation (ENSO) ocean/atmospheric phenomena. El Nino, an oceanic component, is associated with the replacement of the cool upwelling Peruvian coastal current by warmer equatorial waters. The Southern Oscillation, an atmospheric component, is a fluctuation in the intertropical atmospheric circulation, most commonly known as the Walker Circulation. The Southern Oscillation manifests itself as a quasi-periodic (2-4 year) variation in large-scale sea-level pressure, surface wind, and sea-surface temperature anomalies over a wide area of the Pacific Ocean basin (Glantz, 1991).

This research focuses on such oscillations in global SSTs and atmospheric circulation patterns and their effects on the spring and summer severe weather seasons in the southeastern and south-central portions of the United States. Using standard statistical methods of regression and classification trees, this study creates a climatological algorithm for forecasting months ahead, the degree of severity of the spring and summer severe weather seasons for DoD installations within the area of interest.

1.2 Research Objectives

Seasonal forecasts produced using multiple forms of regression and classification tree techniques are at the cutting edge of current weather prediction technology. The goal of this study is to attempt to create a climatological algorithm for use in producing long-range forecasts. This study examines spring and summer severe weather parameters and compares them to SST records and known global circulations from the previous winter season to produce the climatological algorithm, since relationships are found, which are statistically significant.

The specific objectives necessary to achieve the goal of this study were to:

1. define the SST indices, global circulation indices, and severe weather parameters pertinent to the study;
2. identify the regions of interest and examine six stations for an accurate and representative coverage of each region;

3. collect precipitation data from these individual stations. Heavy precipitation was chosen to define severe weather since this data set is most abundant and readily available;
4. gather lightning data within 50 nautical miles of each station. This radius was specifically chosen since weather warnings/watches are issued within it, and previous research has found this radius to be most representative of lightning in the surrounding area of a location;
5. examine tornado data within a 50 nautical mile radius of each station;
6. collect thunderstorm data from each of the six chosen stations;
7. compare the lightning, precipitation, tornado, and thunderstorm data from each station to the global SST indices and the circulation indices using traditional statistical methods of regression;
8. use classification tree techniques to introduce new predictive techniques by combining SSTs and global circulations and explore any relationships worthy of prediction;
9. identify relationships between February and winter indices, regional trends, and prominent global circulation/SST patterns;
10. after detecting if any statistical relationships exist, produce a climatological algorithm for forecasting the intensity of the spring and summer severe weather seasons.

II. Literature Review

2.1 Background on Global Atmospheric Circulations and SSTs Influences

Circulations and currents within the atmosphere and the ocean transport energy from one part of the globe to another. Strong winds force the flow of the surface waters, which results in an upwelling of deep water in certain regions of ocean basins. The combination between this upward convergence cooling surface SSTs and solar heating warming SSTs results in gradients along the ocean surface (Trenberth, 1991). Consequently, the oscillation between the cooling and warming SSTs induces increasing/decreasing pressure gradients over the ocean surface. This change in pressure enhances global circulations and the strength of upper atmospheric winds illustrating the strong interaction between the oceans and the atmosphere (Trenberth, 1991).

Predicting the interaction between the oceans and the atmosphere has been a major challenge for all scientists, however, it has been discerned that global circulations and SSTs play a major role on weather and climate of the world (Gatenbein, 1995). To better understand global circulations, two approaches have been used to obtain temporal correlations: the teleconnection method and the rotated principle component analysis (RPCA). The teleconnection method uses meteorological parameters between one geographical location and correlates them with other point locations in its domain (Barnston, 1987). A teleconnection usually includes two to four centers of action, with the strength of the correlation used to determine whether or not the global circulation is peaking or is of significant strength.

The RPCA uses entire flow field values in a specific region of meteorological parameters to determine where the centers of action are, instead of pre-assigning centers of action like the teleconnection method. This process takes full advantage of large-scale global circulation patterns to produce robust solutions. There are several reasons why RPCA has not been fully used as the primary approach for analysis. Teleconnections are simpler to compute and less removed from the original data, and understanding all aspects of RPCA is difficult because of its interpretability (i.e., what they actually mean physically). However, both methods are analyzed to create indices across the globe.

2.2 The Southern Oscillation Teleconnection Index

One of the most conspicuous of many teleconnections in the world influencing weather and climate is the Southern Oscillation (SO). The evolution of the SO and its corresponding anomalies in pressure have been studied and well documented over the years. The SO refers to the seesaw pattern of atmospheric pressure differences across the tropical Pacific over some time period (Figure 1). An inverse relationship between air pressure in the western Pacific at Darwin, Australia and the south-central Pacific at Tahiti influences major climatic changes across the globe. Interest in the SO increased after 1983, when the 1982-83 ENSO event disrupted global weather patterns making scientists pay closer attention to its corresponding indices (Wagner, 1985). The Southern Oscillation Index (SOI) has been linked to great temperature extremes, flooding, and severe weather and it serves as an efficient predictor for North American weather patterns

(Ting, 1997). The SO index equation that is used by the U.S. Climate Prediction Center (CPC) is defined as:

$$SOI = \frac{\dfrac{Actual\_Tahiti\_SLP - Mean\_Tahiti\_SLP}{Standard\_Deviation\_Tahiti} - \dfrac{Actual\_Darwin\_SLP - Mean\_Darwin\_SLP}{Standard\_Deviation\_Darwin}}{Monthly\_Standard\_Deviation} \quad (1)$$

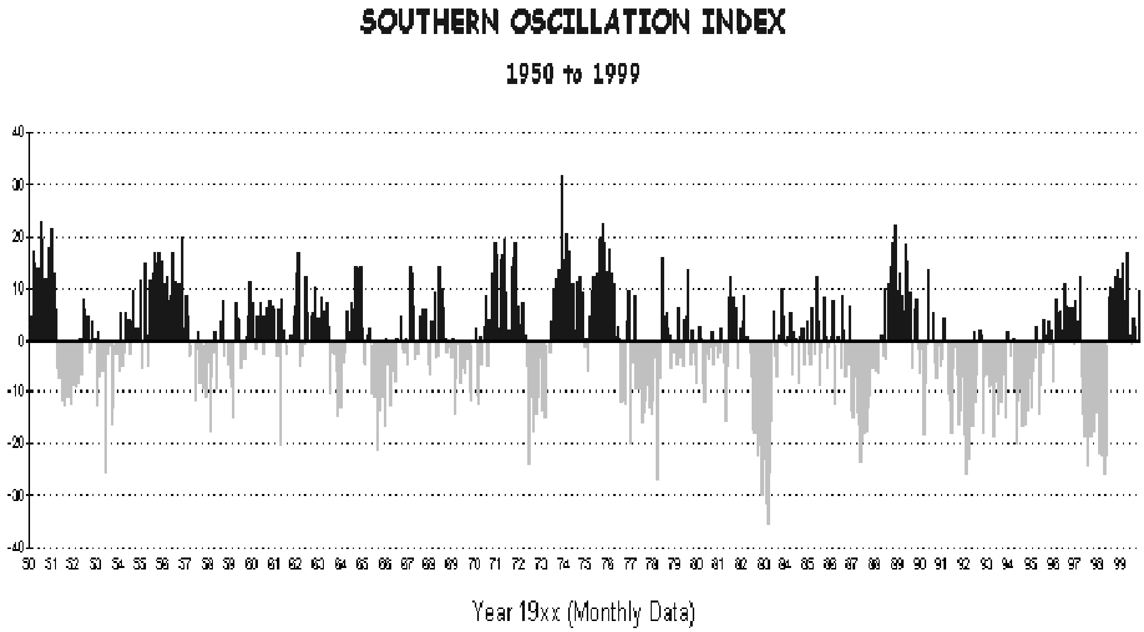

Figure 1. Seesaw pattern of the SOI with a strong, negative phase during the 1982-83 event disrupting global patterns everywhere (Daly, 2001).

2.3 RPCA Indices

The technique for determining other prominent global circulations is RPCA. In this analysis, patterns are determined each month by using specific height anomalies for the three-month period centered on the month. RPCA produces robust indices since it is based on an entire flow field, and not just from height anomalies at specific locations.

The most prominent RPCA global circulation found in all months is the North Atlantic Oscillation (NAO). The NAO correlates part of a strong center over Greenland

with an opposite field over the Atlantic, Europe, or the United States (Figure 2). Research has shown that positive phases of the NAO result in above normal temperatures in the eastern United States and northern Europe, while negative phases produce opposite results. In addition, strong positive phases induce below-normal precipitation over southern Europe. During the mid-1950's though the late 70's, the wintertime NAO showed almost complete domination of the negative phase, and then, a transition to the positive phase until the mid 90's. Thus, the NAO is strongly recognized in winter studies (Hurrell, 1995).

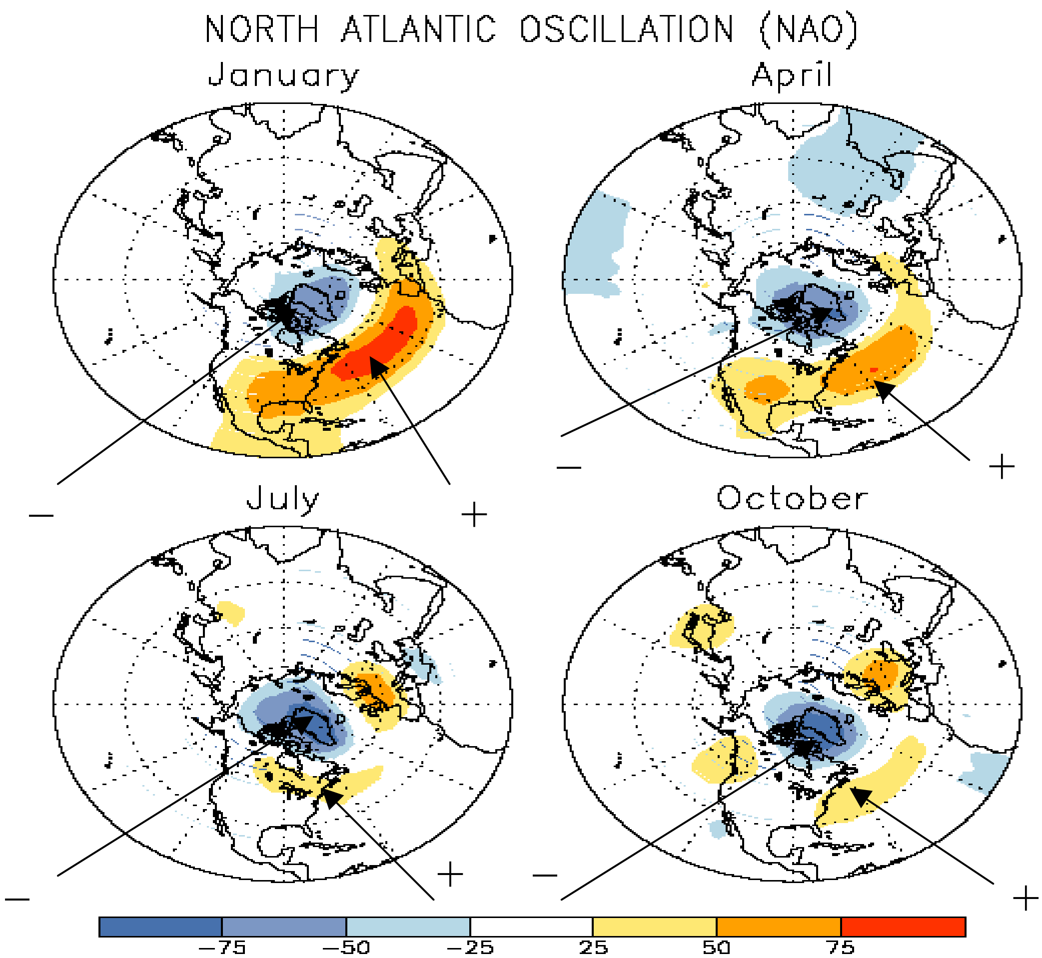

Figure 2. Phases of the NAO with scale of correlation values between the average 700 mb height at a grid point and the RPCA value (U.S. CPC, 2001).

Another prominent global circulation in the Northern Hemisphere is the Pacific/North American (PNA) pattern (Figure 3). The PNA has four strong centers of height anomalies, with two sets of similar signs. The first set is the Aleutian Island height anomaly and the southeastern United States height anomaly. The second set's center is located in the vicinity of Hawaii and near the United States-Canadian border between the Pacific Ocean and Rocky Mountains. Research has shown that the PNA index has encouraging correlations with precipitation. Thus, the PNA pattern is important in the climatic variability in many regions, especially during the winter months when the pattern is a major mode of atmospheric variability (Leathers, 1991).

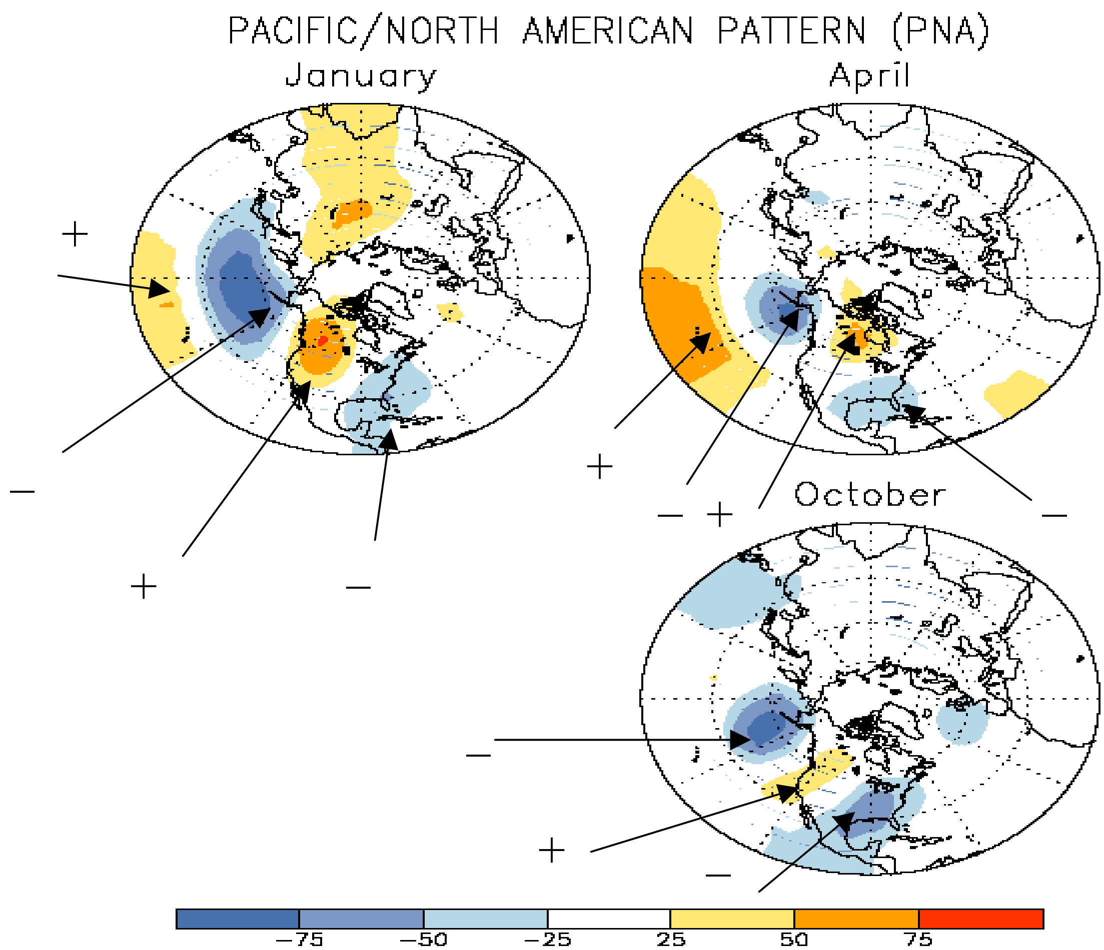

Figure 3. Phases of the PNA with scale of correlation values between the average 700 mb height at a grid point and the RPCA value (U.S. CPC, 2001).

The West Pacific Oscillation (WP) is a global circulation over the North Pacific and appears in all months. During the winter, the pattern orients in a north-south pattern with one center located over the Kamchatka peninsula and another of the opposite sign located in portions of southeastern Asia (Figure 4). In the summer, the WP introduces a third prominent center over Alaska and the Beaufort Sea (Barnston, 1987). The WP moves progressively westward from summer through winter and vice-versa from winter through summer. Due to the wave-like pattern, strong positive or negative phases enhance zonal variations in the location and intensity of the Pacific jet stream, thus becoming a major pattern during the winter (Wallace, 1981).

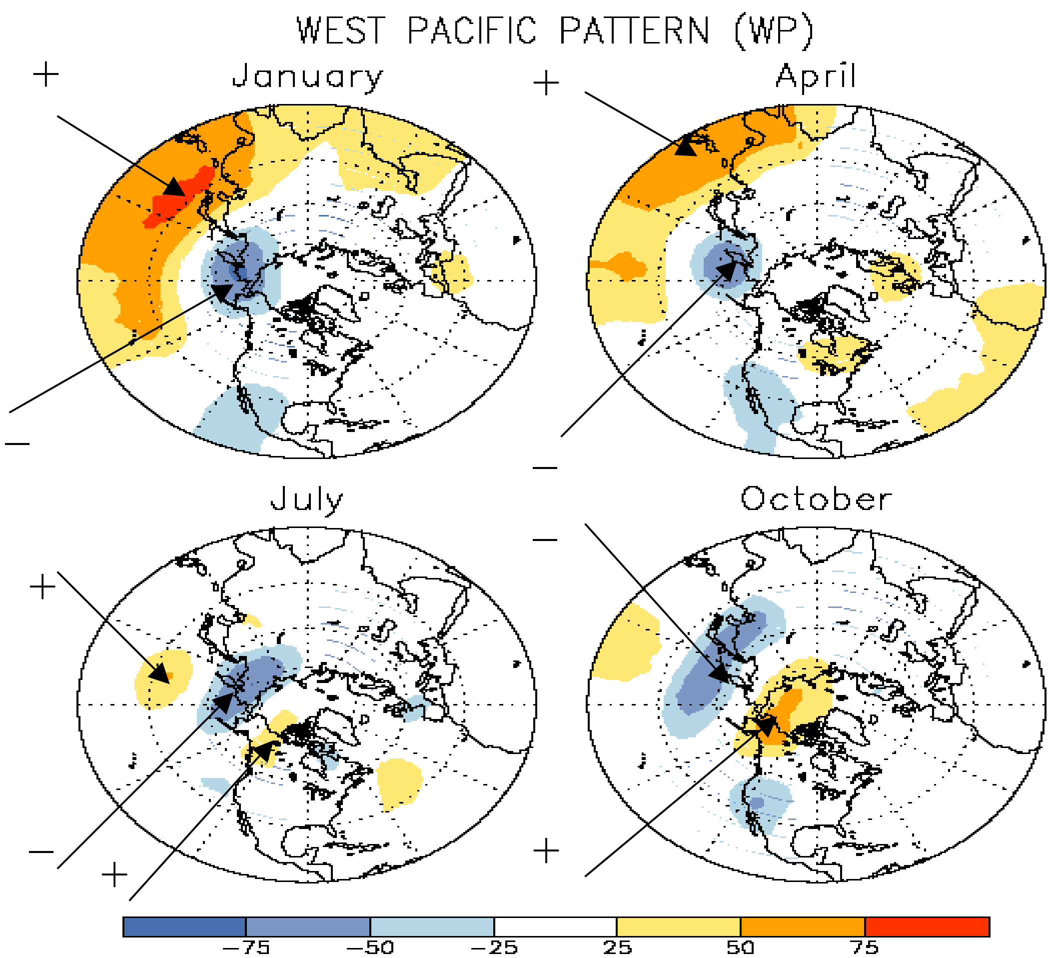

Figure 4. Phases of the WP with scale of correlation values between the average 700 mb height at a grid point and the RPCA value (U.S. CPC, 2001).

Another RPCA global circulation pattern examined is the East Pacific (EP) pattern. A center near Alaska and the west coast of Canada and an opposite sign near Hawaii define it (Figure 5). During positive phases, a deep trough settles over western North America with a pronounced northeastern expansion of the Pacific jet stream. In addition, the subtropical jet stream is generally stronger during this phase and creates above-normal precipitation over the central United States, which brought floods to the Midwest in the summer of 1993. On the other hand, strong negative phases of the EP pattern reduce the intensity through split flow of the jet, creating blocking patterns further east over the Rockies (Barnston, 1987).

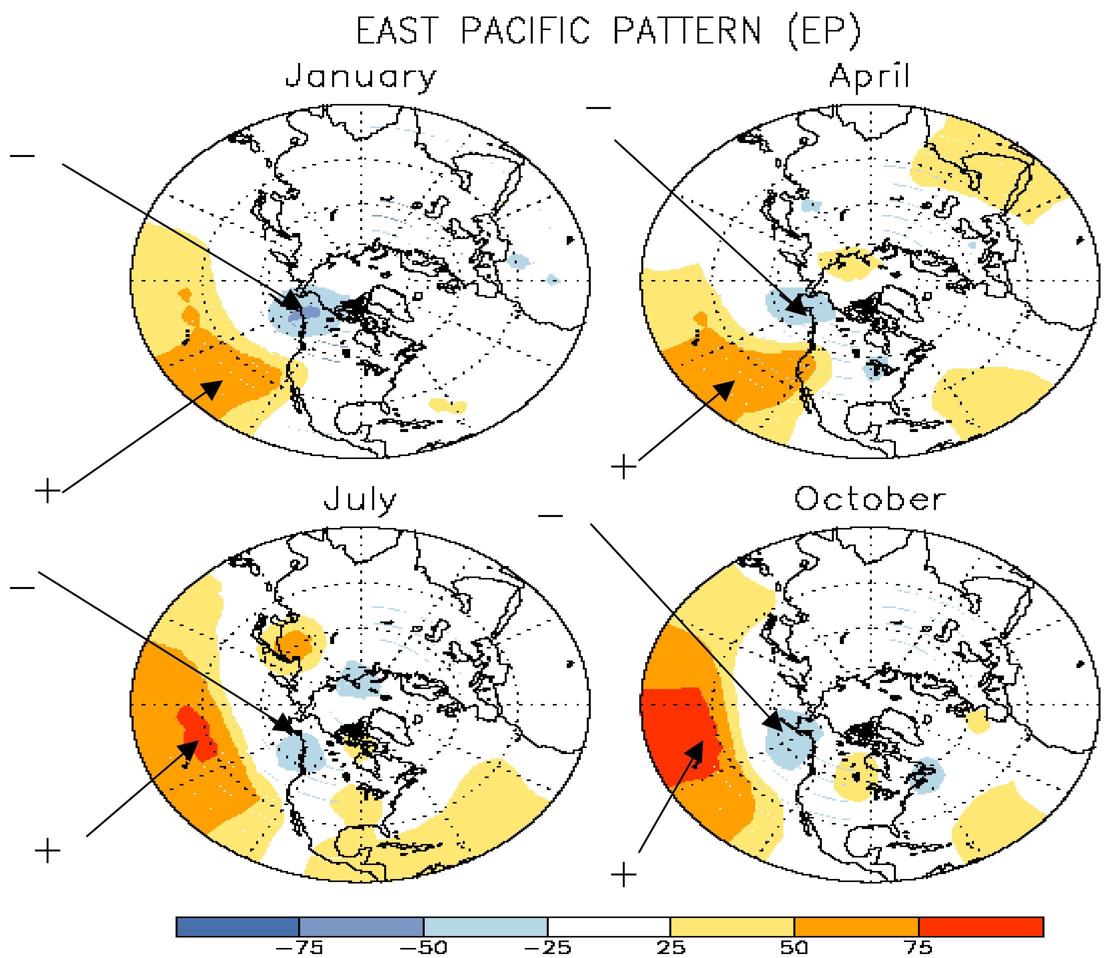

Figure 5. Phases of the EP with scale of correlation values between the average 700 mb height at a grid point and the RPCA value (U.S. CPC, 2001).

The Tropical/Northern Hemisphere (TNH) pattern also strongly influences the polar jet stream and its features are shifted east to be out of phase with the PNA. It has a center just off the Pacific Northwest coast of the United States with a center of the same sign near Cuba (Figure 6). Another center with an opposite sign is located just south of the Hudson Bay (Barnston, 1987). Research has shown that in the winter, when the TNH pattern is in the negative phase, the Pacific jet stream intensifies and its location is shifted well southward into central California (Barnston, 1991). Thus, this global circulation regulates and transports the flow of warmer, marine air and colder, continental air into the United States.

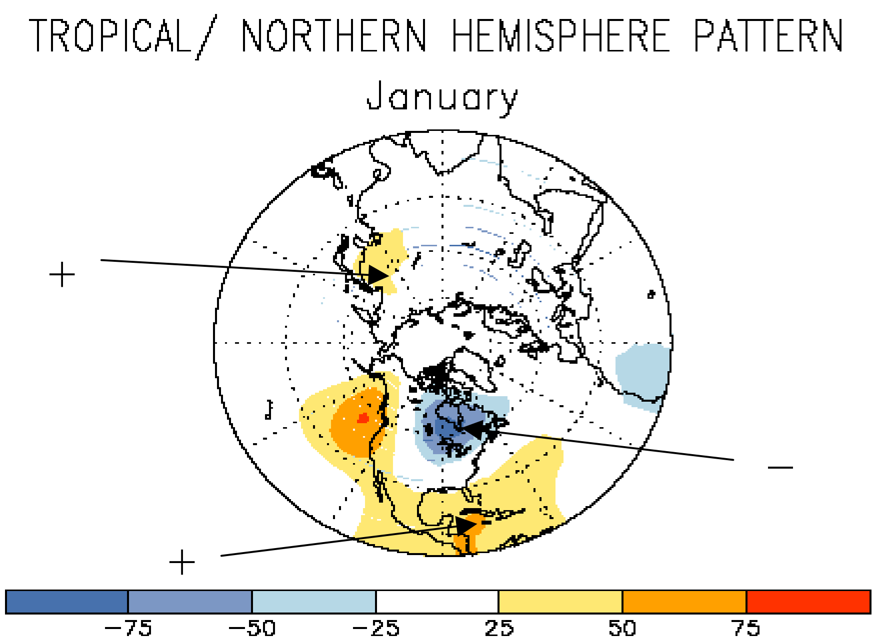

Figure 6. Phases of the TNH with scale of correlation values between the average 700 mb height at a grid point and the RPCA value (U.S. CPC, 2001).

Other well known RPCA indices include the North Pacific pattern (NP), the East Atlantic Jet Pattern (EA-JET), and the Asia Summer pattern (ASU). However, their

significance in the winter months is minimal and will not be introduced since this research is focusing on winter indices used to identify trends with spring and summer severe weather.

2.4 SST Indices

The global circulations that moderate the atmospheric winds link the components of the atmosphere and ocean. Above-normal precipitation over the United States is often associated with excessive moisture transport from the ocean and its associated frequent storm activities passing over the United States. It has been suggested that the primary cause of drought is the change in the atmospheric circulation across North America by changes in SSTs (Trenberth, 1992). SSTs all over the globe are analyzed, and indices are created based on actual SSTs and their respective anomalies. For example, the linkage between Pacific SSTs and United States precipitation was shown to influence the central and eastern United States through the change of atmospheric circulations leading to strong changes in moisture transport (Ting, 1997). Warm SST anomalies in the tropical Pacific have been associated with a decrease in precipitation in North Carolina while cold SST anomalies have shown the opposite results (Roswintiarti, 1998). SSTs have a huge impact globally since the Northern Hemispheric jet streams extract significant amounts of moisture from all oceanic basins. One could ask if this increase or decrease in moisture result in an increase or decrease in severe weather from regimes across the globe, or if there is a balancing effect with the amount of wind shear these jets produce?

Considerable amounts of upper-level wind shear in any thunderstorm event might eventually spell destruction of the storm system itself.

2.5 Severe Weather Parameters

Both global circulations and SSTs have a large but unknown effect on severe weather. The primary variable controlling the enhancement in thunderstorm activity is the position and strength of the jet streams. The increase in southeastern United States thunderstorm activity during the 1997-98 season is directly attributable to the stronger than normal upper-level polar jet stream across the region. Increased baroclinicity associated with the enhanced jet produced a 100-200 percent increase in lightning flashes and lightning days along the Gulf Coast (Goodman, 2000). This increase in the strength of the jet resulted from changing conditions in the Pacific SSTs. However, the underlying feature is that SSTs and global circulations are not directly responsible for the formation of individual thunderstorms, but rather, they are directly related to synoptic flow patterns (Rhome, 2000). In spring 1984, following a strong negative phase of the SO, the United States experienced severe intense storm systems that produced devastating tornadoes. Impacts such as major tornado outbreaks that stretched from Oklahoma to Minnesota and eastward from northern Illinois to Lake Michigan induced F3 and F4 intensities that struck at night causing high casualties and heavy damage. No place on earth is more visited by these storms than the United States. Meteorologists are constantly searching for improved long-range severe weather forecasting techniques.

Their hope is to reduce weather-induced loss of life and property by investigating the interactions between the earth's oceans and atmosphere.

III. Data

The primary objective of this study was to find predictive relationships between global atmospheric circulation and SST indices with certain parameters indicative of the severe weather season in two regions of the United States. In addition, after any predictive relationships are identified, this study created algorithms for forecasters to use based on any strong relationships found. A strong relationship is likely related to regional effects that control the occurrence of severe storms as well as favorable conditions for upper-level forcing mechanisms.

3.1 Regions of Study

Recently, Air Force Weather (AFW) reorganized into regional forecast Hubs across the United States known as operational weather squadrons (OWSs). These OWSs provide meteorological products to aid in the protection of Air Force resources in all military installations in their respective coverage region. This study encompasses two of the four continental Hubs; specifically, the 28th OSW at Shaw AFB and the 26th OWS at Barksdale AFB. Their coverage includes the southeastern and the south-central portion of the United States. Within each OWS area of responsibility (AOR), three bases were chosen for a comprehensive representation of the coverage area (Figure 7).

The southeastern stations chosen were:

1. Shaw AFB, South Carolina
2. Warner-Robins AFB, Georgia

3. Pope AFB, North Carolina.

The south-central stations chosen were:

1. Barksdale AFB, Louisiana
2. Tinker AFB, Oklahoma
3. Randolph AFB, Texas.

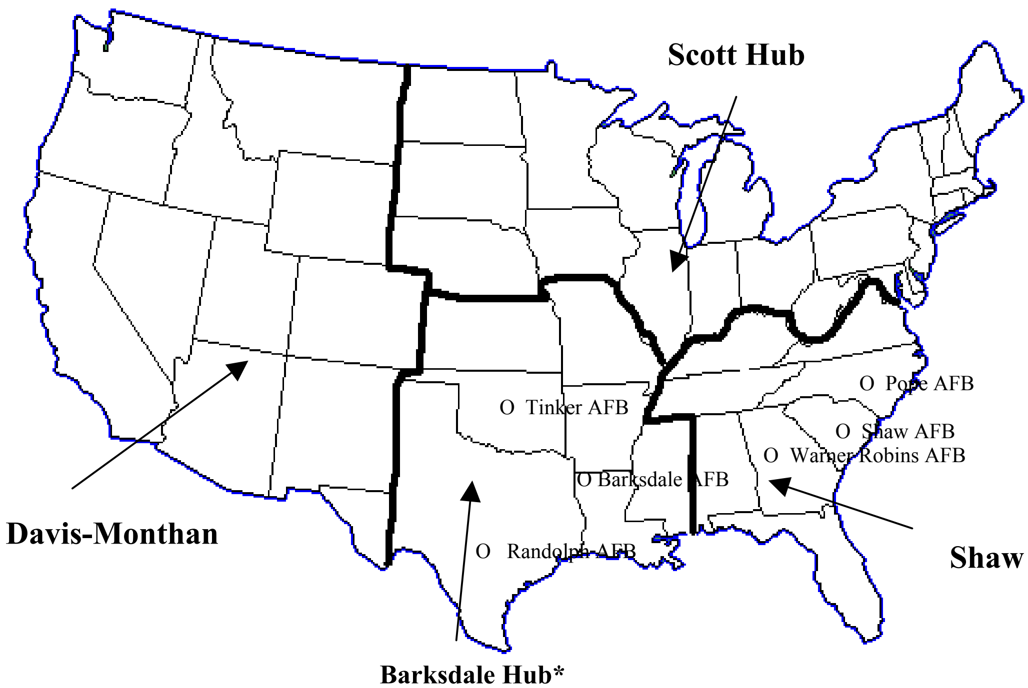

Figure 7. The four Air Force Weather Hubs along with the six stations used in this study (*only two Hubs used in this study).

3.2 Predictors: Teleconnection Index and RPCA Indices

The predictor data in this study are broken up into two sets of variables. The first set is the teleconnection and RPCA indices, which were obtained from the CPC. For all indices except the TNH index, three consecutive monthly values, December through

February were averaged to create a single, winter value. In addition, just the February indices were examined since the averaging of the indices might factor out any trends near the end of the winter season that might prove crucial in finding correlations with the spring and summer severe weather seasons. As there were no February data for the TNH index, the TNH index will not be used in the February only comparisons, therefore, the averaging procedure was applied to the two months of December and January to create the TNH pattern's winter index. Winter values were chosen since these indices are highly significant during the winter season and the goal is to predict the spring and summer severe weather seasons based off of these highly significant winter indices. The indices that were examined are the:

1. Southern Oscillation (SO)
2. North Atlantic Oscillation (NAO)
3. Pacific/North American Pattern (PNA)
4. West Pacific Pattern (WP)
5. East Pacific Pattern (EP)
6. Tropical/Northern Hemisphere Pattern (TNH).

The winter values were examined for each year of the fifty-year period of record (POR), 1951-2000, and compared with the spring and summer severe weather parameters. The fifty-year POR was chosen since such a large data set will stabilize patterns and best identify trends that exist. In addition, data on these indices were readily available from CPC. This is invaluable in any predictive study since the data for any forecast tool developed must be readily available to users. If not, such a tool is only valuable to the researcher themselves.

3.3 Predictors: SST indices

The second set of predictor data includes the SST indices that were also collected from the CPC. Specifically, the SST indices (Figure 8) that this study examined were the:

1. North Atlantic (NATL): 5-20° North, 60-30° West
2. Global Tropics (TROP): 10° South - 10° North, 0-360°
3. Nino 3.4 (NINO): 5° North-5° South, 170-120° West
4. West Coast of United States (WESTUS): Along ship track #1.

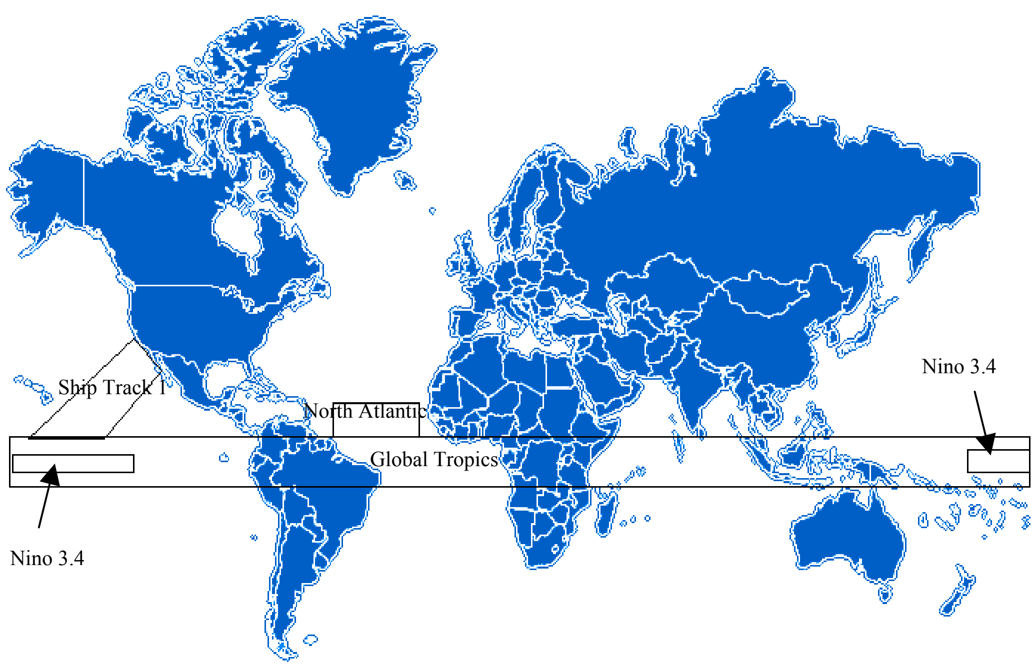

Figure 8. The four SST basins used in this study.

The indices were examined from December through February and averaged over the period to create single, winter values as well as using the February data by themselves. These indices were not anomalies to SSTs, however, since they were the actual mean of the SSTs within their respective ocean basins. Anomalies were not chosen over the actual SST data since this research examined only the winter season of SSTs, therefore using anomalies to factor out the seasonal effects is not necessary. In addition, the winter values were examined each year of the 50-year POR, 1951-2000, and were also compared with the spring and summer severe weather season parameters.

3.4 Predictands: Severe Weather Parameters

The data sets predicted are the severe weather parameters. Each severe local storm season, defined as March though May for spring and June through August for summer, is described by specific parameters. Any of the following parameters were used to illustrate severe weather events:

1. Lightning data within 50 nautical miles
2. Precipitation data greater or equal to 0.50 inches
3. Tornado data within 50 nautical miles
4. Thunderstorm observational data

Lightning data were collected from AFCCC and are analyzed over an 11-year POR, 1990-2000, since accurate coverage was first available at the beginning of the 1990s. The number of lightning days per month was summed for spring and summer to

create single, cumulative values for each season indicative of the total lightning activity within that season.

Precipitation data were calculated from AFCCC and examined over the entire 50-year POR, 1951-2000. The number of days with precipitation greater or equal to 0.50 inches was also summed for the spring and summer seasons to create single, cumulative values for each season. The value of 0.50 inches was chosen over 0.10 inches since this research was examining severe weather events, and while a 0.10 event may have severe weather associated with it, there would also be many events where the 0.10 threshold was met but severe weather had not occurred.

Tornado data were collected from AFCCC and examined over a 45-year POR, 1951-1995. The number of days with tornadoes within 50 nautical miles was also summed for spring and summer to equal a total number of days during the season. Tornado records before the 1980s is questionable, especially since older records relied primarily on observational data alone. With this in mind, tornadoes might be missed at night and in rural areas; therefore, the data presented would represent the minimum number of tornado occurrences.

Finally, thunderstorm data were collected from AFCCC and examined over a 50-year POR, 1951-2000. The number of days with thunderstorms was also summed during the spring and summer seasons to create a single value for each season. Since thunderstorms typically can be heard from 12 nautical miles away, this presents a different data set than the lightning data, and one that has a longer POR that can be used for better regression results. It was anticipated that a relationship exists with at least one

of the parameters, especially, since vast amounts of both predictors and predictand values were analyzed.

IV. Results

4.1 Traditional Statistics

Regression analysis deals with examining relationships between two or more variables. The simplest mathematical relationship between two variables is the linear relationship:

$$y = B_o + B_1 \cdot x + \varepsilon \tag{2}$$

In this case, the predictand is the y-value and the predictor is the x-value (introduced in chapter 3). B_o represents the y-intercept parameter while B_1 represents the slope of the line parameter. These parameters are determined by using the method of least squares fit. The method of least squares fit minimizes the sum of squared distances from each point to the line that best fits. Since this study focuses on multiple predictors, global circulations and SSTs, multiple linear regression was used. In multiple linear regression, the simple linear regression model is adjusted just by adding on the extra predictors. The general additive multiple linear regression equation is:

$$y = B_o + B_1 \cdot x_1 + B_2 \cdot x_2 + ... + B_k \cdot x_k + \varepsilon \tag{3}$$

In this equation, k is the number of predictors used for each model. For this study, k will be nine for the Feb indices (excluding TNH) and 10 for the winter indices. Multiple linear regression also uses the method of least squares fit and is the method of choice to perform traditional statistics.

4.1.1 Methodology

Before any regression can occur, 20% of the data should be excluded from any tests for uses of model verification. If a valid model does exist, then the excluded data can be used to verify model accuracy. Since this data uses sample sizes near 50 (number of years), 10 years have to be excluded for the optimal 20% verification. The 10 years that were removed using a random number generator are: 1956, 1957, 1967, 1974, 1978, 1982, 1987, 1990, 1997, and 2000. In addition to excluding data, data sets need to be checked to determine whether they are continuous or discrete. Since precipitation >0.50 inches, thunderstorms, and lightning events are numerous during the spring and summer seasons in the southeastern and south-central United States, these data sets don't have any problems with being a continual data set. However, since tornadoes are not frequent, especially for most of the east coast, tornado data are discrete and will not be included in the standard regression process.

After data was excluded for verification purposes and checked for being continual, a regression model was created including all predictors into the equation. For significance to occur in any model, the p-value must be lower than the standard alpha level of 0.05. The p-value is the last number located in the Analysis of Variance (ANOVA) table under the F Ratio column. A p-value less than 0.05 indicates that the model does fit better than simply the mean. Individual predictor p-values can be checked in the parameter estimates table shown above under the Prob>t column. For an even more efficient model, these individual p-values can be examined and excluded to increase

the significance of the model, and eventually the adjusted coefficient of determination (R^2), similar to the process within stepwise regression.

Once significance of the model has been achieved, the coefficient of determination was checked to account for the total variation in the predictand (y-value) explained by all the predictors (x-values). R^2 values range from 0 to 1, and if there was no linear relationship between the predictand and predictors, R^2 is 0 or very small. If all observations fall on the best fit line, R^2 is 1. However, the estimate of R^2 tends to be rather optimistic of the population, therefore adjusted R^2 was used to more closely reflect how well the model fits the population and is usually more analyzed for models with more than one predictor.

When using regression analysis, problems such as multicollinearity occur. Sometimes in regression analysis, there was a close relationship between two or more predictors, which results in high errors for the parameter estimates. When multicollinearity may be a problem within the model, the variance inflation factor (VIF) was checked. Any predictors with multicollinearity problems have large variance inflation factors. Severe VIFs include any value over 20. If any severe instances occur, the correlation matrix between predictors will be analyzed to see how strong the relationship exists between the predictors. The model will be reanalyzed and one of the predictors with a higher adjusted R^2 and a lower individual p-value will be kept in the model, while the predictor with the lower adjusted R^2 and a higher individual p-value will be discarded.

In addition to problems with multicollinearity, influential data points are also checked and removed to make a more efficient model. With smaller samples such as the

lightning data set, influential data points occur often. Since this problem was drastic and hard to overcome with such small sample sets, the lightning data was excluded for all regression processes. With the larger sample sets, such as precipitation and thunderstorms, influential data points are not an issue.

4.1.2 Analysis

Once multicollinearity and influential data points are satisfied, the model was in its polished form. Only coefficients of determination with significant, p-values <0.05 found in the ANOVA table, are listed in Table 1 and Table 2 below, otherwise, no sig. appears.

Table 1. Adjusted R^2 between spring/summer thunderstorm days & Feb/winter indices.

Region	Station	Spring vs Feb	Spring vs Winter	Summer vs Feb	Summer vs Winter
Southeast	Shaw	no sig	no sig	0.107	no sig
	Pope	0.175	no sig	0.276	0.271
	Robins	no sig	no sig	no sig	no sig
South-central	Barksdale	0.089	0.087	0.297	0.193
	Randolph	0.219	0.414	0.352	0.150
	Tinker	0.234	0.189	0.104	no sig

Table 2. Adjusted R^2 between spring/summer precipitation days & Feb/winter indices.

Region	Station	Spring vs Feb	Spring vs Winter	Summer vs Feb	Summer vs Winter
Southeast	Shaw	0.144	0.274	0.254	0.133
	Pope	0.177	0.307	0.093	no sig
	Robins	0.330	0.257	no sig	no sig
South-central	Barksdale	no sig	0.205	0.421	0.287
	Randolph	no sig	0.262	no sig	no sig
	Tinker	no sig	no sig	0.161	no sig

Finding R^2 between spring/summer severe weather parameters and Feb/winter indices was the key focus for multiple linear regression. In addition, differences between the Feb and winter indices, southeast and south-central regions, and global circulations and SSTs were examined. Overall, R^2 values ranged from about 0.10-0.40, which are all rather weak correlations for uses in prediction, therefore no model was created to help with the final algorithm. However, knowing that correlations do exist proves valuable uses in statistics and show that the indices do show some sign of relationship with precipitation >0.50 and thunderstorm events.

Another goal of this study was to determine whether averaging all the winter months into one value would show better correlations than just looking at the end of the season trend. With averaging, the entire season was included into the process, although specific events, especially near the end of the season are not taken into full account. The advantage of just looking at February indices would show how the atmosphere along with oceanic processes are changing to possibly identify trends and patterns with the upcoming spring and summer severe weather season. After analyzing Table 1, equally weak correlations existed between spring vs. Feb indices and spring vs. winter, however, more correlations existed with Feb indices in the summer months than the winter indices. Looking at Table 2, equally weak correlations existed between spring vs. Feb and spring vs. winter, however, more correlations existed with winter indices in the spring than the Feb indices. Factoring in both Table 1 and Table 2, there seems to be no apparent advantage of using Feb indices over an averaged winter index, since even though Feb indices proved to show more relationships with precipitation >0.50 data, winter indices showed more relationships with the thunderstorm data.

The next goal of multiple regression was to identify if any regional trends existed. To accomplish this, a trend was identified if a global circulation or SST pattern was significant, p-value <0.10 (a more lenient p-value), in all three stations in their respective region. The only regional trend that was identified was the spring precipitation vs. the winter indices model run. Both the PNA and the NATL indices correlated with all three stations in the southeast, although the correlations were weak. Since the PNA does have a center of action over the southeast and the NATL is close in proximity to the southeast region, the indices that were closer to the regions of interest did have more significance in the regression models.

Finally, the last goal considered during multiple linear regression was to determine whether global circulations of SST patterns appeared more frequently in the models. Table 3 shows the number of occurrences that an index was significant, <0.10, in any model run. The results show that the NATL appeared most frequent followed by NINO. Nineteen signals were identified by NATL and NINO identified 15 signals, and overall, SSTs showed more relationship with severe weather than the global circulations.

Table 3. Number of occurrances that an index was significant (<.10) in Feb/winter

Model	SO	NAO	PNA	WP	EP	TNH*	NATL	TROP	NINO	W US
Spring Thunderstorm	0	4	3	1	2	0	5	1	2	2
Summer Thunderstorm	1	1	3	2	4	1	4	4	6	3
Spring Precipitation	3	2	3	5	2	2	5	4	5	5
Summer Precipitation	2	4	3	1	3	1	5	2	2	2
Total	6	11	12	9	11	4	19	11	15	12

*lower values for TNH since no winter model run

Overall, even though R^2 values were weak (<0.50) for all model runs, statistical conclusions can be drawn from the analysis. First, there was no apparent advantage of looking at February indices over winter indices, however, this process was used again for data mining and regression trees since the data are already formatted and deeper relationships could have been overlooked. Second, the proximity of an index to the region will increase the significance and eventually the correlation of the model. Both the PNA and the NATL had greater influence on the southeastern region than other indices. Finally, multiple linear regression showed that SST indices appeared more often in the model runs than did global circulations. Even though R^2 remained low, the results above provided helpful information in the data mining and regression tree processes. Knowing what key indices to use for each model would aid in the tree building process and eventually into an algorithm usable by OWS forecasters.

4.2 Classification and Regression Trees (CART)

CART analysis deals with complex relationships involving several predictands and predictors, and was used in this research when traditional statistics had been exhausted. From the thunderstorm, precipitation, and tornado data sets, CART established classification trees that predicted a categorical predictand. These classification trees consist of binary decision rules that split nodes (decision points) either to the left or right based on a test against a significant predictive value and will continue to branch until a terminal node (final node) was reached (Burrows, 1992). CART

provided a way to examine data and discover important grouping cases to formulate rules and to make predictions. The key elements of the CART analysis are:

1. choosing the best splitting technique for the trees,
2. designing the trees for the best predictive results,
3. validating the tree through cross-validation techniques.

4.2.1 Methodology

CART works by choosing a split at each node so that each child node was more pure than its parent node. In a completely pure node, all of the cases have the same value for the categorical, target variable. CART defaults the measure of the split impurity using the Gini splitting rule. Gini looks for the largest class in the database and strives to isolate it from other classes. For example, if the minimum node number of cases was set to 5, nodes with total sample size of 4 or less will not split, however, nodes with total sample size of 5 or more will continue to split once the threshold value of 5 was met. After initial splitting in the tree was made, the process was repeated until the most pure terminal nodes are reached. While this approach may seem short sighted since it attempts to separate classes by focusing on one class at a time, Gini performance is frequently so precise and is considered the best splitting rule.

The next key element of the CART analysis was designing a tree for the best predictive results. The most pure terminal nodes in a tree will have 100% of the data formulated into one category, therefore if all the criteria were met to arrive at that terminal point in that specific tree, 100% of the time that specific category will be

predicted. CART also provided a misclassification matrix to show risk estimates. The risk estimate is the proportion of cases correctly classified that indicates the extent to which the tree makes accurate predictions. If a tree was completely pure, the actual category would match up with the predicted category and the risk estimate would be zero. This might seem like the ideal tree, however it still does not provide any insight into validation of the tree. Therefore, the 10-fold cross validation technique was used for validation. The combination of a pure terminal node for 100% predictability and a low cross-validation risk estimate would provide for the best design of a tree.

The final key element of the CART analysis was validating the tree. There are several methods of validation, however, the 10-fold cross-validation method was used in this study since it is an improvement over the traditional holdout method, where a certain percent is removed from the data, when dealing with a smaller sample size. Since this study deals with sample sizes of 50 or less, removing data using the holdout method will only decrease the sample size more and a robust validation will not be achieved. The 10-fold cross validation is a method for estimating what the error rate of 10 sub-trees would be if there was test data. The optimal tree, which was derived from the first two key elements, was tested using 10 subsets. After the data were divided into 10 subsets, one of the 10 subsets was used as the test set and the other 9 subsets are put together to form the training set. Then the average error across the 10 trials was computed. The advantage of this method is it does not matter how the data gets divided, and that the variance of the resulting estimate is reduced as the number of folds is increased. Evidence has been shown that using 10-20 folds gives better results than a smaller number.

In this study, obtaining the most pure terminal nodes and the lowest cross-validation risk estimate was done by rerunning several trees, each with different splitting thresholds. Usually a splitting threshold of two would create trees without impurities, however, the cross-validation risk estimate could be higher. When a splitting threshold of five was used, the tree would have impurities, however the cross-validation risk estimate could be lower. Finding the perfect balance between the lower impurities and the lower cross-validation was the main challenge during the analysis.

4.2.2 Analysis

Before any classification trees could be created, the thunderstorm, precipitation, and tornado data sets had to be categorized to best solve the problem to this research. Just like the tradition statistics portion of the research, lightning data wasn't used during the CART analysis due to the small size of the data set. The goal was to answer how intense the severe weather season would be, and a classification into below normal, normal, and above normal categories was achieved through ranking the data into equal thirds. However, since all data sets contained seasonal values, the data couldn't be split exactly into equal thirds, although for the thunderstorm and precipitation data sets, the data was split close enough to fit into the below normal, normal, and above normal categories. Tornado data proved more of a challenge. Since the data wasn't normally distributed, which was a problem during traditional statistics, not all the data could be split into equal thirds after ranking the data occurred, therefore, some of the tornado data was split into equal thirds, while other data sets were split 50%/25%/25%. These splits

were determined to be the climatology of the data sets, which was shown in the result tables further in this research. The goal of the classification trees would be to improve upon the climatology determined by the splits above.

After ranking and splitting the data into below normal, normal, and above normal categories, the classification trees were created (Appendix A). The next step was to determine if the tree was the best tree for creating an algorithm for forecasters to use. In order to determine if the best tree was created, several factors had to be determined:

1. the purity of the tree,
2. the sample size of the terminal nodes,
3. the cross-validation risk estimate.

All of these factors were used to reach the improvement over climatology, which only was shown in the results if it was better than 0%. First, the purity of the tree was determined. Only terminal nodes of 100% were used to obtain the highest improvement. Terminal nodes less than 100% were not chosen since the cross-validation risk estimate multiplied by any terminal node less than 100% would not result in any improvement above climatology.

Next, any terminal node sample size less than three would not be used since two years of data did not represent at least 5% of the thunderstorm and precipitation data sets. This same process was used for continuity in the tornado data sets.

Finally, obtaining the lowest cross-validation risk estimate was achieved by rerunning trees with different stopping rules explained in the CART methodology section of this research. Subtracting the cross-validation risk estimate from 100% would result in the tree accuracy. Once the tree accuracy was determined, the difference from

climatology was determined by subtracting the tree accuracy from the climatology. Then, the improvement over climatology would be that difference divided by the climatology. Once all improvements were shown to be above 0%, the criteria were used as determined from the tree to provide a forecast algorithm to predict the intensity of each severe weather category.

4.2.3 CART Results

Result tables were broken up regionally to identify trends with the global circulation and SST indices. Since the goal was to obtain the best forecast accuracies for the algorithm, February indices and winter indices were both used to create trees, however, only the best index was shown and is shown in the criteria with capitalized indices being the winter indices and lower-case indices being the February indices. If the criterion were met for either the February or winter indices, a long-range forecast would provide for the intensity, either below normal, normal, or above normal, and a forecast accuracy for the algorithm.

Table 4 results show the southeast spring thunderstorm forecast algorithm. The best forecast accuracies were for Pope AFB with 47% accuracies-a 42% improvement over climatology. The best regional trend identified was the SO index, which was signaled in every station for use in predicting spring thunderstorms in the southeast region. Both winter and February indices were used to provide the best forecast algorithm for southeast spring thunderstorms.

Table 4. Southeast spring thunderstorm forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Below Average	natl≤25.70 25.3<nino≤27.30 ep>0.20 wpo≤-0.20	42% / 33% / 27%
	Average	natl≤25.70 nino≤25.30	42% / 33% / 27%
	Average	natl>25.70 ep≤0.75 nino>26.40	42% / 33% / 27%
	Above Average	natl≤25.70 25.30<nino≤27.30 ep≤0.20 so>-1.10	42% / 33% / 27%
Pope	Below Average	so>-0.95 0.50<ep≤1.35	47% / 33% / 42%
	Average	so≤-0.95	47% / 33% / 42%
	Above Average	so>-0.95 ep≤0.15 nao>-.05 natl≤25.60	47% / 33% / 42%
Robins	Below Average	PNA>0.83 NAO>-0.10	43% / 33% / 30%
	Below Average	-0.42<PNA≤0.83 SO>-0.70 TROP≤27.41 WESTUS>22.10 WPO≤.48	43% / 33% / 30%
	Average	PNA≤0.83 SO≤-0.70 TNH≤0.20 NAO≤0.85	43% / 33% / 30%

*winter indices are capitalized

Table 5 results show the south-central spring thunderstorm forecast algorithm. A 46% tree accuracy was acknowledged for Randolph AFB-a 39% improvement over climatology. Regional trends identified were the NATL, EP, and PNA indices. They were all signaled for predicting spring thunderstorms in the south-central region. Both winter and February indices were used to provide the best algorithm for south-central spring thunderstorms.

Table 6 results show the southeast summer thunderstorm forecast algorithm. Up to 45% tree accuracies were acknowledged for Shaw AFB-a 36% improvement over climatology. No stations had the best predictive results for both summer and spring thunderstorms, and no regional trends were identified for predicting summer thunderstorms in the southeast region, however only winter indices were used to provide the best algorithm for southeast summer thunderstorms.

Table 7 results show the south-central summer thunderstorm forecast algorithm. A 48% tree accuracy was acknowledged for Randolph AFB-a 45% improvement over climatology. In addition, Randolph AFB continually had the best predictive results for both summer and spring thunderstorms. NAO was the only signal identified in all stations in the south-central region for predicting summer thunderstorms. Both February and winter indices were used to provide the best algorithm for south-central spring thunderstorms.

Table 5. South-central spring thunderstorm forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average	$NATL>26.40$ $EP\le 0.83$	43% / 33% / 30%
	Below Average	$NATL\le 26.40$ $0.01<NAO\le 0.71$ $-0.72<PNA\le 0.07$	43% / 33% / 30%
	Average	$NATL\le 26.40$ $NAO\le -0.36$ $TROP>27.50$	43% / 33% / 30%
	Average	$NATL\le 26.40$ $NAO>-0.36$ $PNA>-0.07$ $WPO\le -0.25$ $TNH\le 0.95$	43% / 33% / 30%
	Above Average	$NATL\le 26.40$ $NAO>0.71$ $-0.55<PNA\le -0.07$	43% / 33% /30%
	Above Average	$NATL\le 26.40$ $NAO>-0.36$ $PNA>-0.07$ $-0.25<WPO\le 0.95$	43% / 33% / 30%
Randolph	Below Average	$EP>-0.50$ $NATL>25.90$ $SO>-0.70$ $NAO\le 1.25$	46% / 33% / 39%
	Average	$EP>-0.50$ $NATL>25.90$ $SO\le -0.70$ $WESTUS\le 23.30$	46% / 33% / 39%
	Average	$EP>-0.50$ $NATL\le 25.90$ $PNA\le 0.07$ $NAO>0.04$	46% / 33% / 39%
	Above Average	$EP\le -0.50$ $NATL\le 26.00$	46% / 33% / 39%
Tinker	Below Average	$pna>-1.15$ $natl>25.90$ $wpo>-0.10$	43% / 33% / 30%
	Average	$pna>-1.15$ $natl>25.50$ $wpo\le -0.10$	43% / 33% / 30%
	Above Average	$pna>-1.15$ $natl\le 25.50$ $ep\le -0.15$ $westus\le 25.80$	43% / 33% / 30%

*winter indices are capitalized

Table 6. Southeast summer thunderstorm forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Below Average	WPO≤0.39 TROP≤27.41 TNH>0.30 NAO≤1.13	45% / 33% / 36%
	Average	WPO≤0.39 27.4<TROP≤27.5 EP≤0.46	45% / 33% / 36%
	Average	WPO>0.39 PNA>0.48 SO>3.15	45% / 33% / 36%
	Above Average	WPO≤0.39 TROP>27.50 TNH>0.05	45% / 33% / 36%
Pope	Average	25.7<NATL≤26.30 -0.29<PNA≤0.72 NAO>-0.48 WPO≤0.46	37% / 33% / 12%
Robins	Below Average	EP≤-0.32 TROP≤27.70 SO≤0.73	39% / 33% / 18%
	Average	EP>-0.19 NAO>0.09 SO>-0.24	39% / 33% / 18%

*winter indices are capitalized

Table 7. South-central summer thunderstorm forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average	wpo>-0.75 natl>25.80 nao≤0.55	42% / 33% / 27%
	Below Average	-0.75<wpo≤-0.50 natl≤25.80	42% / 33% / 27%
	Above Average	wpo≤-0.75 natl≤25.40	42% / 33% / 27%
Randolph	Below Average	trop≤28.10 nino>26.60 nao≤-0.05 -1.05<so≤0.25	48% / 33% / 45%
	Below Average	trop≤28.10 nino>26.60 nao>-0.05 wpo>-0.95 westus≤25.70	48% / 33% / 45%
	Average	trop≤27.60 nino≤26.60 ep>-0.95 nao≤-0.20 so≤1.30	48% / 33% / 45%
	Average	trop≤28.10 nino>26.60 nao≤-0.05 so≤-1.05	48% / 33% / 45%
	Above Average	trop≤27.60 25.2<nino≤26.2 ep>-0.95 nao>-0.20	48% / 33% / 45%
	Above Average	27.6<trop≤28.10 nino≤26.60 ep≤1.35	48% / 33% / 45%
	Above Average	trop>28.10	48% / 33% / 45%
Tinker	Below Average	PNA>1.02	44% / 33% / 33%
	Below Average	PNA≤1.02 WESTUS≤22.90 NAO>0.02	44% / 33% / 33%
	Below Average	PNA≤1.02 WESTUS≤22.90 NAO≤0.02 EP≤-0.15 TNH>-0.04	44% / 33% / 33%
	Below Average	PNA≤1.02 WESTUS>23.30 NINO>26.60 SO>-1.10 WPO≤0.65	44% / 33% / 33%
	Average	PNA≤1.02 22<WESTUS≤23 NINO>26.60 SO>-1.10	44% / 33% / 33%
	Above Average	PNA≤0.56 WESTUS>22.90 NINO>26.60 SO≤-1.10	44% / 33% / 33%

*winter indices are capitalized

Table 8 results show the southeast spring precipitation forecast algorithm. A 57% tree accuracy was acknowledged for Robins AFB-a 73% improvement over climatology. The EP and SO were the only signals identified in all stations in the southeast region for predicting spring precipitation >0.50. Both February and winter indices were used to provide the best algorithm for southeast spring precipitation >0.50.

Table 9 results show the south-central spring precipitation forecast algorithm. A 44% tree accuracy was acknowledged for both Barksdale and Randolph AFB-a 33% improvement over climatology. The EP and NAO were the two signals identified in all stations in the south-central region for predicting spring precipitation >0.50. Only winter indices were used to provide the best algorithm for south-central spring precipitation >0.50.

Table 10 results show the southeast summer precipitation forecast algorithm. A 47% tree accuracy was acknowledged for Robins AFB-a 42% improvement over climatology. In addition, Robins AFB continually had the best predictive results for both summer and spring precipitation >0.50. The WESTUS was the only signal identified in all stations in the southeast region for predicting summer precipitation >0.50.

Table 11 results show the south-central precipitation forecast algorithm. A 45% tree accuracy was acknowledged for Barksdale AFB-a 36% improvement over climatology. In addition, Barksdale AFB continually had the best predictive results for both summer and spring precipitation >0.50. The EP was the only signal identified in all stations in the south-central region for predicting summer precipitation >0.50. Both February and winter indices were used to provide the best algorithm.

Table 8. Southeast spring precipitation >0.50 forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Below Average	PNA>-0.35 EP≤0.27 TROP≤27.50	40% / 33% / 21%
	Below Average	PNA>-0.25 EP≤0.10 TROP>27.50 TNH≤0.75 WPO>-0.27	40% / 33% / 21%
	Below Average	-0.35<PNA≤-0.07 EP>0.27 TROP>27.10	40% / 33% / 21%
	Average	PNA>-0.07 EP>0.27	40% / 33% / 21%
	Average	PNA≤-0.35 SO>0.39 NAO>-0.62	40% / 33% / 21%
	Above Average	PNA≤-0.35 SO≤0.39 WPO≤1.06	40% / 33% / 21%
Pope	Below Average	EP>-0.73 -1.06<SO≤-0.14 NINO≤26.40	42% / 33% / 27%
	Average	EP>-0.73 SO>-1.06 NINO26.40	42% / 33% / 27%
	Average	EP>-0.73 SO≤1.06 NATL≤25.97 NAO≤0.79	42% / 33% / 27%
	Average	-0.73<EP≤0.12 SO>-0.14 NINO≤26.40 WESTUS>21.90	42% / 33% / 27%
	Average	EP>0.12 SO>0.29 24.9<NINO≤26.4 WESTUS>21.88 TNH>-0.20	42% / 33% / 27%
	Above Average	EP≤-0.73 NAO≤0.75	42% / 33% / 27%
	Above Average	EP>-0.73 SO≤-1.06 NATL>25.97	42% / 33% / 27%
Robins	Below Average	wpo≤-1.05 ep≤1.05	57% / 33% / 73%
	Average	wpo>-1.05 -0.55<pna≤1.10 trop≤27.93 nao≤1.55 ep≤1.35	57% / 33% / 73%
	Above Average	wpo>-1.05 pna≤-0.55 trop≤27.72 so≤1.35	57% / 33% / 73%

*winter indices are capitalized

Table 9. South-central spring precipitation >0.50 forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average	EP>-0.22 WESTUS>23.30 NAO≤0.53 NATL≤26.49	44% / 33% / 33%
	Average	EP>-0.50 WESTUS>23.16 0.53<NAO≤0.96	44% / 33% / 33%
	Average	EP≤-0.50 PNA>0.39	44% / 33% / 33%
	Average	EP>-0.50 WESTUS≤23.10 NAO≤1.22 TROP≤27.65 NATL>25.60 WPO>-1.33	44% / 33% / 33%
	Above Average	EP≤-0.50 PNA≤0.39 NAO>-1.15	44% / 33% / 33%
Randolph	Below Average	-0.50<EP≤0.17 NATL>26.30	44% / 33% / 33%
	Below Average	0.06<EP≤1.15 25.8<NATL≤26.3 NAO≤0.96 PNA>-1.08	44% / 33% / 33%
	Average	-0.50<EP≤0.07 NATL≤26.30	44% / 33% / 33%
	Average	EP>0.07 NATL≤25.81 PNA≤-0.19	44% / 33% / 33%
	Above Average	-0.93<EP≤-0.50 NAO≤0.89	44% / 33% / 33%
	Above Average	EP>0.07 NATL≤25.80 PNA>-0.19 NAO≤0.61	44% / 33% / 33%
Tinker	Below Average	WESTUS≤23.05 NAO≤-0.44	42% / 33% / 27%
	Average	WESTUS<22.48 EP>0.61	42% / 33% / 27%
	Average	WESTUS>23.05 NAO≤0.79 TNH>-1.22 WPO>-0.22	42% / 33% / 27%
	Above Average	22.5<WESTUS≤23 NAO>-0.44	42% / 33% / 27%

*winter indices are capitalized

Table 10. Southeast summer precipitation>0.50 forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Below Average	natl>25.75 trop≤27.95	45% / 33% / 36%
	Average	natl≤25.75 westus≤24.45 nao>-0.30	45% / 33% / 36%
	Above Average	natl≤25.75 westus>24.45 -1.10<pna≤0.75 nao≤-0.05 so>-1.30	45% / 33% / 36%
Pope	Below Average	PNA>-1.06 TROP≤27.70 SO≤0.19 NINO≤27.13 WPO>-0.32	37% / 33% / 12%
	Average	PNA>-1.06 TROP>27.70 SO>-3.15 WESTUS≤24.47	37% / 33% / 12%
	Average	-1.06<PNA≤0.80 TROP≤27.70 0.19<SO≤1.62 EP>-0.53 NINO>24.86	37% / 33% / 12%
Robins	Below Average	25.4<natl≤25.9 ep≤-0.25 nao>-1.45 wpo>-1.00	47% / 33% / 42%
	Average	natl>25.43 ep>-0.25 so≤0.60	47% / 33% / 42%
	Average	natl≤25.43 westus>24.65 -0.70<nao≤0.50	47% / 33% / 42%
	Above Average	natl≤25.43 westus≤24.65 nino>25.24 pna>-1.35	47% / 33% / 42%

*winter indices are capitalized

Table 11. South-central summer precipitation >0.50 forecast algorithm.

Station	Category	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average	nao>-0.05 natl>25.34 0.15<ep≤1.80	45% / 33% / 36%
	Average	-1.25<nao≤-0.05 natl≤25.72 trop≤27.75	45% / 33% / 36%
	Above Average	nao>0.05 natl>25.34 -0.60<ep≤0.15 so>-1.25	45% / 33% / 36%
	Above Average	0.05<nao≤0.80 natl≤25.34	45% / 33% / 36%
Randolph	Below Average	SO>-1.32 WPO≤0.19 26.0<NATL≤26.7 PNA≤0.83 EP>-0.60	43% / 33% / 30%
	Average	SO>-1.32 0.19<WPO≤0.55	43% / 33% / 30%
	Average	SO>-1.32 WPO≤0.19 25.8<NATL≤26.0 PNA≤0.72	43% / 33% / 30%
	Above Average	SO≤-1.32 WPO≤0.85	43% / 33% / 30%
Tinker	Below Average	PNA≤0.75 WESTUS>21.98 WPO>-0.35 -0.25<NAO≤0.99 NATL>25.86 EP≤1.12	42% / 33% / 27%

*winter indices are capitalized

Table 12 results show the southeast spring tornado forecast algorithm. A 49% forecast accuracy was acknowledged for Robins AFB-a 96% improvement over climatology. The NAO and PNA were the only signals identified in all stations in the southeast region for predicting spring tornadoes. Both February and winter indices were used to provide the best algorithm for southeast spring tornadoes.

Table 13 results show the south-central spring tornado forecast algorithm. A 47% forecast accuracy was acknowledged for Barksdale AFB-a 42% improvement over climatology. The SO was the only signal identified in all stations in the south-central region for predicting spring tornadoes. Only winter indices were used to provide the best algorithm for south-central spring tornadoes.

Table 14 results show the southeast summer tornado forecast algorithm. A 47% forecast accuracy was acknowledged for Pope AFB-a 42% improvement over climatology.   The WPO, EP, and NAO were the signals identified in all stations in the southeast region for predicting summer tornadoes. Only winter indices were used to provide the best algorithm for southeast summer tornadoes.

Table 15 results show the south-central tornado forecast algorithm. A 58% forecast accuracy was acknowledged for Randolph AFB-a 132% improvement over climatology was noted.   The TROP was the only signal identified in all stations in the south-central region for predicting summer tornadoes. Both February and winter indices were used to provide the best algorithm for south-central summer tornadoes.

Table 12. Southeast spring tornado forecast algorithm.

Station	Category (# of tornadoes)	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Average (1)	NAO≤-0.02 PNA>-0.90 TROP≤27.74	45% / 33% / 36%
	Above Average(>1)	NAO>-0.02 WESTUS≤22.95 SO>-0.25	45% / 33% / 36%
Pope	Below Average (0)	PNA>-0.46 TNH>-0.55 WESTUS≤23.13 TROP>26.81 NAO≤0.82 NATL≤25.60	44% / 33% / 33%
	Below Average (0)	PNA>0.85 TNH≤-0.55 NATL≤26.37 WPO≤0.95	44% / 33% / 33%
	Below Average (0)	PNA>-0.46 TNH≤-0.55 NATL>26.37	44% / 33% / 33%
	Average (1)	-0.46<PNA≤0.85 TNH≤-0.55 NATL≤26.37	44% / 33% / 33%
	Average (1)	PNA≤-0.46 WPO>-0.70 TNH≤1.20	44% / 33% / 33%
	Above Average (>1)	PNA>-0.46 TNH>-0.55 WESTUS>23.13 EP≤0.25	44% / 33% / 33%
	Above Average (>1)	PNA≤-0.46 WPO≤-0.70 NAO≤0.38	44% / 33% / 33%
Robins	Below Average (0)	25.13<natl≤23.36 nao≤-0.30	49% / 25% / 96%
	Above Average (>2)	25.13<natl≤25.36 nao>-0.30 wpo≤-0.50	49% / 25% / 96%

*winter indices are capitalized

Table 13. South-central spring tornado forecast algorithm.

Station	Category (# of tornadoes)	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average (0-1)	NAO≤-0.12 TROP≤27.41 TNH≤0.88	47% / 33% / 42%
	Below Average (0-1)	NAO>-0.12 WESTUS>23.45 PNA≤0.72 SO>-1.25	47% / 33% / 42%
	Average (2-3)	NAO>-0.12 WESTUS≤23.45 NATL≤25.75	47% / 33% / 42%
	Above Average(>3)	NAO≤-0.12 TROP>27.41 TNH>-0.72	47% / 33% / 42%
Randolph	Average (1)	-0.40<WPO≤0.95 SO≥-0.70 NATL≤26.07 PNA>-0.71	40% / 33% / 21%
	Above Average (>1)	WPO>-0.40 SO>-0.70 TROP≤27.70 TNH>-0.85 NINO>24.88	40% / 33% / 21%
Tinker	Below Average (1-2)	EP≤-0.30 -0.33<WPO≤0.47	40% / 33% / 21%
	Above Average (>4)	EP≤0.85 WPO≤-0.33 PNA≤1.00 SO≤0.44	40% / 33% / 21%

*winter indices are capitalized

Table 14. Southeast summer tornado forecast algorithm.

Station	Category (# of tornadoes)	Criteria*	Tree Accuracy / Climatology / Improvement
Shaw	Below Average (0)	WESTUS≤22.64 WPO≤-0.42	44% / 33% / 33%
	Below Average (0)	23<WESTUS≤23.5 WPO≤-0.42 EP>-0.80	44% / 33% / 33%
	Average (1)	WESTUS≤23.16 WPO>0.19 NAO>-0.67	44% / 33% / 33%
	Average (1)	WESTUS>23.45 TNH>-0.78	44% / 33% / 33%
	Above Average (>1)	22.6<WESTUS≤23 WPO≤0.02 TNH≤0.95	44% / 33% / 33%
	Above Average(>1)	WESTUS>23.45 TNH≤-0.78 NAO>-0.20	44% / 33% / 33%
Pope	Below Average (0)	WESTUS>22.78 EP>0.12 PNA≤0.73 NAO>-1.02	47% / 33% / 42%
	Average (1)	22<WESTUS≤22.8 -0.73<NAO≤0.71	47% / 33% / 42%
	Average (1)	WESTUS>22.78 EP≤0.12 SO≤-0.04 0.27<WPO≤1.06	47% / 33% / 42%
	Above Average (>1)	WESTUS>22.78 EP≤-0.62 SO≤-0.04 WPO≤0.27	47% / 33% / 42%
Robins	Average (1)	WPO≤-0.08 PNA>-0.85 EP>0.53	35% / 25% / 40%
	Average (1)	WPO≤0.19 PNA>-0.85 EP≤-0.33 NAO≤-0.25	35% / 25% / 40%
	Above Average (>1)	WPO≤0.19 PNA≤-0.85	35% / 25% / 40%
	Above Average (>1)	WPO≤0.19 PNA>-0.30 -0.33<EP≤0.53 NAO≤-0.25	35% / 25% / 40%

*winter indices are capitalized

Table 15. South-central summer tornado forecast algorithm.

Station	Category (# of tornadoes)	Criteria*	Tree Accuracy / Climatology / Improvement
Barksdale	Below Average (0)	-0.55<PNA≤0.18 EP≤0.92	51% / 50% / 2%
	Average (1)	PNA>0.48 WPO≤0.67 SO≤0.15	51% / 25% / 104%
	Average (1)	PNA≤-0.55 NAO>-0.62 TROP≤27.40	51% / 25% / 104%
Randolph	Below Average (0)	0.20<ep≤1.20	58% / 50% / 16%
	Below Average (0)	ep≤0.20 27.8<trop≤28.1	58% / 50% / 16%
	Below Average (0)	ep≤-0.45 trop≤27.72	58% / 50% / 16%
	Average (1)	ep>1.20 nao>0.05	58% / 25% / 132%
	Above Average (>1)	-0.45<ep≤0.20 trop≤27.82 wpo>-0.45	58% / 25% / 132%
Tinker	Below Average (0-1)	NATL≤26.06 WESTUS≤23.48 TROP>27.16 PNA>0.86	56% / 33% / 70%
	Average (2)	NATL≤26.06 WESTUS>23.48 NAO>0.11 WPO≤0.82	56% / 33% / 70%
	Above Average (>2)	NATL≤26.06 WESTUS≤23.48 TROP≤27.16 EP≤0.75 NAO>0.32	56% / 33% / 70%
	Above Average (>2)	NATL>26.06 WESTUS≤23.35 SO>1.20	56% / 33% / 70%

*winter indices are capitalized

If the criteria were not met at all, then climatology would still be the best prediction, however, there was a significant increase in the algorithm over climatology using all three severe weather parameters. Since the three weather parameters are dependent sets with each other, it would be difficult to combine the three data sets into one severe weather product, and a lot of information would be lost in the combination process. The advantage of keeping the data sets individualized was that specific long-range forecasts could still be made with each severe weather parameter. In addition, the three severe weather parameters only partially define the severe weather season since there are other parameters that could be used to define it at as well. Therefore, the algorithms in the tables above are to be used separately to characterize the severe weather season.

Regional trends within the algorithms were difficult to recognize, however, connections between indices and the severe weather parameters were made. The EP index was noted several times with the south-central spring and summer precipitation forecasts, and the NAO was noted several times with the southeast spring and summer tornado forecasts. However, no further research was done on these findings since that would have been another major path that would have swayed from the goal of this research.

Other trends were also recognized from the results. Randolph AFB continually had the best predictive results for both seasonal thunderstorm forecasts within the respective region. Robins AFB and Barksdale AFB continually had the best predictive results for both precipitation forecasts within their respective regions.

Overall the CART results were positive. They confirmed that algorithms with reasonable predictability could be produced for forecasting the intensity of the severe weather season. The predictive tables produced in this study are deemed ready to use by AFCCC and OWS forecasters to answer such questions each year.

V. Conclusions and Recommendations

5.1 Conclusions

The main goal of this research was to create a climatological algorithm if statistical relationships were found between spring and summer severe weather parameters and SST and global circulation indices. Forecast algorithms were created using CART analysis, specifically classification trees, which improved upon climatology on multiple cases. Thunderstorm data showed improvements up to 45%. Precipitation data showed improvements up to 73%. Finally, tornado data showed improvements up to 132%. The specific objectives (stated in Chapter 1) were all met to design the predictive algorithms.

SST indices, global circulation indices, and severe weather parameters were all defined. Global circulation indices were divided into two categories: teleconnection and RPCA. Both categorical indices were used and the results show that both types had influences on severe weather parameters, however, the RPCA provided robust indices because of an encompassing spatial domain. The severe weather parameters, thunderstorm, precipitation >0.50, tornado, and lightning data, were used to define the spring and summer severe weather seasons. Lightning data would have been used in all statistical approaches, however, the small sample size (10 years) created severe limitations (Objective 1).

The identified regions of interest were the southeast and south-central portions of the United States. Accurate representation of each region was adequately covered with

three stations in each region. The three stations provided insight into certain climatological spatial trends that existed within each region (Objective 2).

Thunderstorm, precipitation >0.50, tornado, and lightning data were all collected and readily available from AFCCC. Limitations did exist with all data sources and should not be forgotten when analyzing the results, however, a larger sample size was used, except lightning data, to help eliminate the effects from these limitations. During the CART analysis, these severe weather parameters were ranked and categorized in the classification tree process (Objectives 3-6).

After data were collected, thunderstorm, precipitation >0.50, and tornado data from each station were compared to the global SST and circulation indices using traditional statistical methods of regression. Overall, R^2 values were weak (<0.50) for all model runs, however, prominent statistical conclusions were pulled from the analysis. Proximity of an index to the region of study was noted as a key factor for a high significance within the model. In addition, multiple linear regression showed that SST indices appeared more often in model runs than global circulations. Understanding the traditional statistical methods did provide insight into the CART analysis (Objective 7).

CART analysis was used once traditional statistics could not design the predictive algorithm. Specifically, classification trees developed forecast algorithms with accuracies better than climatology. If the criteria were not met in any of the algorithms, climatology would still be the best prediction. The three weather parameters were not combined to produce one severe weather product, however, the thunderstorm, precipitation >0.50, and tornado data remained individualized since all three parameters should be used to completely define the severe weather season. Finally, CART analysis,

in addition to traditional statistics, provided conclusions into regional trends identified in this study (Objective 8).

CART analysis and traditional statistics provided conclusions about each data set as well as regional trends. First, they showed that there was no advantage of using February indices over winter indices, therefore, both indices were used in the final classification tree process and climatological, forecast algorithm. Second, the regional trends identified in traditional statistics showed that the PNA and NATL indices correlated well with the three stations in the southeast. Finally, CART analysis showed that the EP showed the best relationship several times with the south-central spring and summer precipitation forecasts, and the NAO showed the best relationship several times with the southeast spring and summer tornado forecasts (Objective 9).

Overall, CART results identified positive trends that existed between the severe weather parameters and the SST and global circulation indices. The thunderstorm data showed improvements up to 45%, the precipitation data showed improvements up to 73%, and finally, the tornado data showed improvements up to 132%. CART confirmed that climatological, predictive algorithms could be produced for forecasting the intensity of the severe weather season (Objective 10).

5.2 Recommendations

There are several limitations and recommendations that should be considered when using such climatological, predictive algorithms. They are as follows:

1. extend the research to examine all global SST and circulation indices. Only the prominent, winter indices were used in this research;
2. acquire more stations within each region to better understand spatial trends and provide forecast algorithms for all stations within the Hub AOR;
3. use lightning data in the statistical process when more years become available. Lightning data provides a more comprehensive coverage of surrounding regions of a station and is less prone to error than thunderstorm data;
4. examine all four Air Force Weather Conus Hubs. The two Hubs examined were the Shaw and Barksdale Hub since past research has shown more relationships between severe weather and global circulation indices in those regions;
5. introduce regressional trees from the CART analysis to create actual forecast numbers or ranges;
6. produce a program that would automatically generate the forecast intensity from the predictive algorithms. As of now, forecasters have to use these algorithms manually, and automation is needed since it would save forecasters time.

Appendix A: Example Classification Tree

This example tree (Figure A) will illustrate the three key factors in creating a classification tree for predictive purposes. This specific tree shows spring thunderstorm data (predictand) at Barksdale AFB compared with all February SST and global circulation indices (predictors). In each node, three categories were analyzed with category 0 being below normal, category 1 being normal, and category 2 being above normal. At node 0, original parent node, the total amount of data is shown (50 in this case) and the three categories. Although the three categories are not split exactly into equal thirds, it is assumed close enough for climatological forecast purposes.

The purity of the tree was determined at each terminal/child node. Only the nodes with 100% were analyzed and used in the algorithms. The nodes that fit this case are node 4, node 7, node 9, and node 15.

Finally, the cross-validation risk estimate would be incorporated to figure out the final forecast accuracy for each node. CART analysis provided the cross-validation risk estimate, and in this case, the error was 60%. Since the error was 60%, then the tree accuracy would be 40%. The improvement would be the tree accuracy minus the climatology divided by the climatology, in this case, 21%.

Any nodes that improved upon climatology (33% in this case) would have shown up in the results, and then their criteria would be recorded into the final predictive algorithm. Since only two nodes proved worthy of the final algorithm in this example, more classification trees, including all winter indices, would have been created to encompass more predictive years.

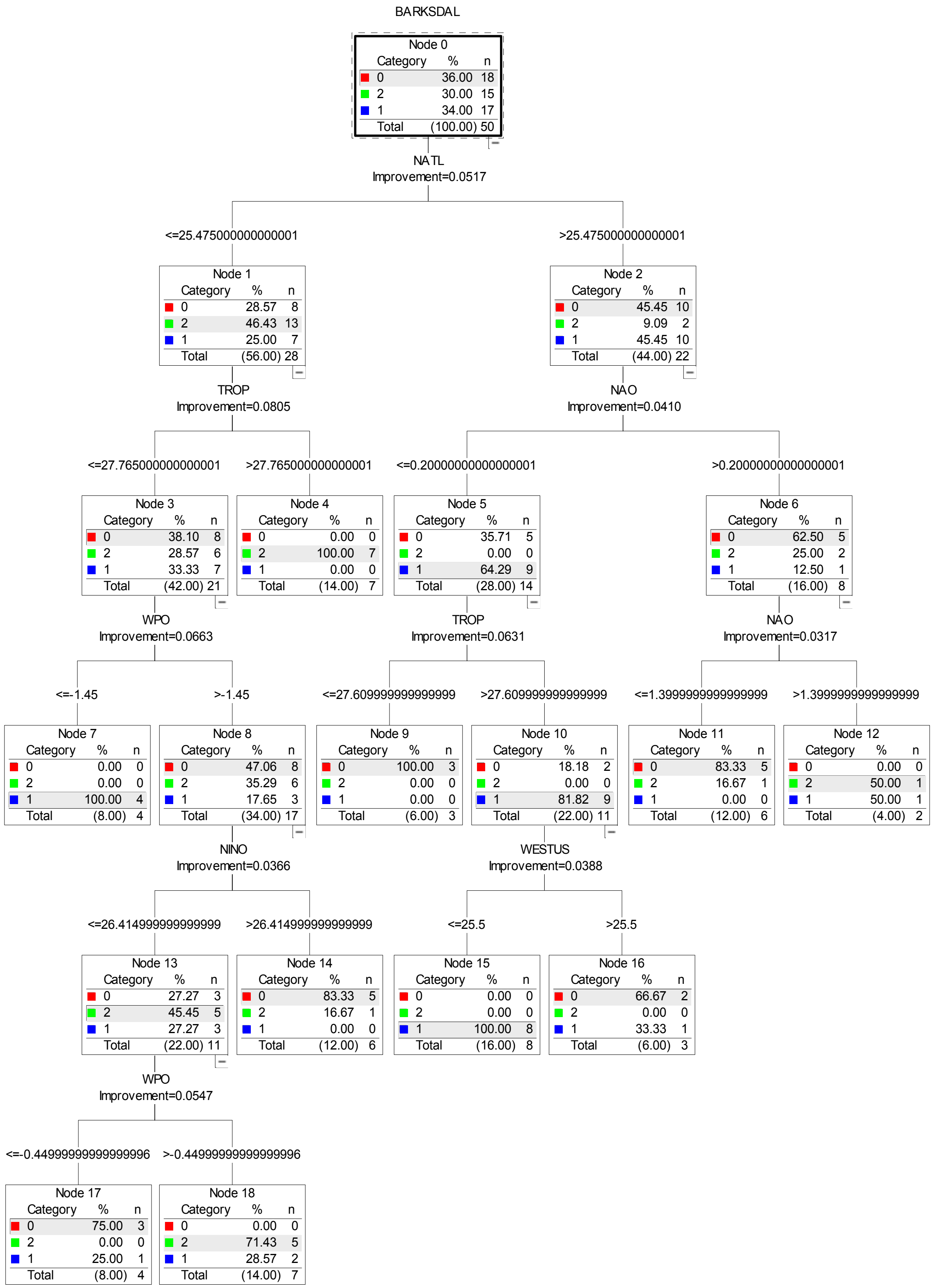

Figure A. An example classification tree that shows spring thunderstorm data at Barksdale AFB compared with all February SST/global circulation indices.

Bibliography

Barnston, G., Anthony., Livezey R.E., 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115, 1083-1126.

Burrows, William R. and Assel, Raymond A. "Use of CART for Diagnostic and Prediction Problems in the Atmospheric Sciences," American Meteorological Society: 12th Conference on Probability and Statistics in the Atmospheric Sciences, 161-166, 1992.

CPC, 2001: http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.html.
Daly, John L., 2001. "The El Nino Southern Oscillation (ENSO)": http://www.microtech.com.au/daly/elnino.htm

Gantenbein, D., 1995: El Nino: the weathermaker. *Popular Science*, 246, 76-82.

Glantz, M.H., 1991: Introduction, *Teleconnections Linking Worldwide Climate Anomalies*, M.H. Glantz, R.W. Katz, & N. Nicholls, Ed., Cambridge University Press, New York, 1-3.

Goodman, S.J., et. al., 2000: The 1997-98 El Nino event and related wintertime lightning variations in the southeastern United States. *Geophysical Research Letters*, 27, 541-544.

Hurrell, J.W.,1995: Decadal trends in the North Atlantic oscillation regional temperatures and precipitation, *Science*, 269, 676-679.

Leathers, D.J., et. al., 1991: The Pacific/North American Teleconnection Pattern and United States Climate. Part I: Regional Temperature and Precipitation Associations. *Journal of Climate*, 4, 517-528.

Rhome, J.R., et. al., 2000: Mesoclimatic analysis of severe weather and ENSO interactions in North Carolina. *Geophysical Research Letters*, 27, 2269-2273.

Roswintiarti, O., et. al., 1998: Teleconnections between tropical Pacific sea surface temperature anomalies and North Carolina precipitation anomalies during El Nino events. *Geophysical Research Letters*, 25, 4201-4204.

Sanders, T., 1985: *Weather. A User's Guide to the Atmosphere*. Icarus Press, Indiana.

Ting, Mingfang., H. Wang, 1997: Summertime U.S. precipitation variability and its relation to Pacific sea surface temperature. *Journal of Climate*, 10, 1853-1873.

Trenberth, K.E., 1991: General Characteristics, *Teleconnections Linking Worldwide Climate Anomalies*, M.H. Glantz, R.W. Katz, & N. Nicholls, Ed., Cambridge University Press, New York, 1-3.

Trenberth, K.E., G.W. Branstrator, 1992: Issues in Establishing Causes of the 1988 Drought over North America. *Journal of Climate*, 5, 159-172.

Wagner, A., James., 1985: Seasonal climate summary, the climate of Spring 1984—An unusually cold and stormy season over much of the United States. *Monthly Weather Review*, 113, 149-169.

Wallace, J.M., and D.S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, 109, 784-812.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
14-01-2002	Master's Thesis	Jun 2001 - Mar 2002

4. TITLE AND SUBTITLE

DESIGNING AN ALGORITHM TO PREDICT THE INTENSITY OF THE SEVERE WEATHER SEASON

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Freestrom, Hugh, J., 1st Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 P Street, Building 640
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GM/ENP/02M-4

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFCCC/DOC5
Maj Joseph King
151 Patton Ave Rm 120
Asheville, NC 28801-5002

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Examination of atmospheric and oceanic circulations may explain interannual climate variability in the Northern Hemisphere on a seasonal scale. It is crucial to develop more accurate seasonal climate forecasts using both global circulation and sea surface temperature indices to aid in long-range weather forecasts. These indices are becoming increasingly available to worldwide users and using them for seasonal prediction has spread to the Department of Defense (DoD). DoD is extremely interested in long-range seasonal forecasts of severe weather for asset protection and worldwide operations. The goal of this research was to create a predictive algorithm for the southeastern and south-central portion of the United States in support of the Air Force Combat Climatology Center (AFCCC) to use in predicting the intensity of the spring and summer severe weather seasons.
Beginning with multiple linear regression, this study found minimal relationships between several severe weather parameters and these known global circulation and SST indices. Although, after little predictive value was found using traditional statistics, CART analyses were developed to create an algorithm for DoD forecasters to use for seasonal severe weather prediction. Results confirmed that algorithms with reasonable predictability can be produced for forecasting the intensity of the severe weather season.

15. SUBJECT TERMS
Atmospheric circulations, Climate, Long-range forecast, Prediction, Multiple linear regression, Classification Trees, Sea-surface temperatures, Algorithms, Severe Weather

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Ronald P. Lowther, Lt Col, ENP
U	U	U	UU	70	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4546